Know the Earth...Show the Way

Metadata Monthly

NGA/DCGS Metadata Harmonization

Issue 38
Nov 2010

Enterprise Sharing: A Process to Expose Data Assets

The Department of Defense (DoD) Military Services are working to achieve the vision of Directive 8320.02, "Data Sharing in a Net-Centric Department of Defense."[1] It calls for the collective use of standard metadata and vocabulary mapping for effective information exchange by "speaking the same language." To move beyond a point-to-point information exchange, DoD data consumers and data providers must transit intermediary hubs that understand both sides of an exchange. This article focuses on the data providers' tasks and describes a bottoms-up implementation process to expose data assets and vocabularies.

Current data providers often organize into communities-of-interest (COIs) that manage their Authoritative Data Sources (ADS) and data sharing practices within that community. To meet Directive 8320.02, data sharing across COIs must also be enabled. The end goal is to prepare data assets to be advertised and accessed across multiple COIs using metadata according to standard XML schemas. The road to prepare for data exposure begins with building the COI vocabulary infrastructure, through COI members posting vocabularies on the DoD Metadata Registry (MDR),[2] and ends with the provisioning of metadata tags for machine-to-machine access. The process is described and illustrated in Figure 1.

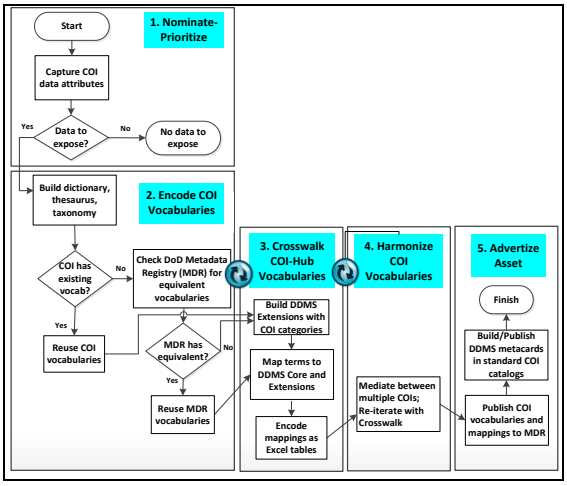

Figure 1. Data Provider Process to Implement Data Sharing Across the Enterprise

Nominate-Prioritize

The process begins with building the list of information asset types for exposure by each ADS. The ADS can be a web site, or applications with executable machine interfaces, relational databases, collections of structured (e.g., text, imagery, and video) or unstructured documents. COI knowledge engineers shall conduct meetings on regular basis with the stakeholders (data stewards and members of the governance bodies), and document agreements in spreadsheets for asset attributes and three main parameters: priority schedule, data values and sharing-readiness. The priority schedule and data values are driven by enterprise sharing needs outside individual COI users; the sharing readiness captures current operational states of the data (e.g., structured vs. unstructured; stand-alone vs. aggregated product; and classification level). The minimum output of this step shall be a table of COI data asset types along with their attributes.

[1] http://www.dtic.mil/whs/directives/corres/pdf/832002p.pdf

[2] https://metadata.ces.mil/mdr/ accessed 11/4/10.

Know the Earth...Show the Way

Metadata Monthly

NGA/DCGS Metadata Harmonization

Issue 38
Nov 2010

Encode-Crosswalk

These steps cover capturing COI terms and connecting them with relational links, thus enabling search beyond string matching. When documents for existing structured assets, applications and databases are exploited for existing metadata, the DoD Discovery Metadata Specifications (DDMS)[3] should be used for guiding selection of appropriate terms. Next, COI data architects shall interview domain experts and capture requirements to build templates for documenting important COI terms and term relationships. During this step, COI data providers shall begin searching the enterprise resources, such as MDR and the DoD Air Force *OneSource*,[4] for re-usable terms. Step 2 is iterative with the Crosswalk step below, and requires human experts to guide encoding COI vocabularies. The minimum output of this step shall be a list of COI-specific tags and controlled vocabularies.

Crosswalk-Harmonize

In these steps, metadata working groups, with representative members from all COIs perform technical mediation to create agreed formats (e.g., XML simple type vs. complex type) and procedures to map multiple COIs to the same hub. This step is often iterative and may require individual COIs to revise their encoding. The minimum output of this step is a list of COI-to-DDMS mappings approved under a governance framework. Additional benefits could also result from mapping COI vocabularies to a hub taxonomy.

Advertise

This step finalizes the infrastructure to share COI vocabularies and mappings to the enterprise by posting them on enterprise platforms such as MDR and *OneSource*. These web-enabled tools let the users share and re-use vocabularies and mappings in an open-market environment. The tools also provide multiple, open-standard vocabularies to build vocabularies and mappings of COI vocabularies. This level of visibility provides an "early detection" for enterprise-wide adoption.

After advertising the asset types and controlled vocabularies, data providers also must provide access to the actual asset instances by integrating the COI data stores to distribution hubs. There are at least three integration types to handle the disparate nature of legacy COI systems, from the simplest wrap-and-reuse interfaces to the most rigorous caching catalog. The three integration types are described in details below. For simplicity, only the vocabulary mapping aspect is discussed, ignoring in this instance important aspects such as security and COTS integration constraints.

Data Access

Applications Interfaces. To access current COI data, standard web services (WS) can wrap existing application interfaces with software components that are DDMS-compliant. For example, the COI user interfaces of a current DoD intelligence processing system[5] query application programming interfaces (API) called the Middleware Data Layer. New query parameters describe the "what, who, where, when:" for example, DDMS title, identifier, description, creator, geo-locations and date stamp. But actual translated values must adapt to the re-used API. The query response shall require translation back from DDMS tags to the consumer "language." Exposing these web services, and their required parameters can occur

[3] http://metadata.dod.mil/mdr/irs/DDMS/ accessed 11/4/10.
[4] https://gcic.af.mil/oneSource/ accessed 11/4/10.
[5] http://www.globalsecurity.org/intell/systems/gccs-i3.htm accessed 11/4/10.

by either a direct URL endpoint or a dynamic lookup registry.[6] This level of data exposure is the least intrusive and simply requires integration of DDMS-compliant application adapters for bidirectional translation.

Data Wrapper. To access current COI data, standard WS can query COI data stores directly with DDMS-compliant tags and return results populated with COI asset metadata. The difference with the simple application interfaces method above is in re-factoring multiple interfaces into a data mapping layer that can be used for several related queries across disparate COIs. The overall benefit comes from a central DDMS translator rather than one for every single application interface. Besides mapping COI metadata to DDMS tags, providers should also encode the mappings as standard XML documents and advertise them on *OneSource* and DoD repositories, including *forge.mil.*[7] This enables re-use of the software components built to process these mappings.

Enterprise Metadata Catalog. This integration level requires the most up front effort. COI experts need to guide the population of COI vocabularies to mapping templates for automated DDMS metacard population before asset ingest. This catalog technique, when combined with resource adapters for both metadata population and query translation, can provide faster query performance and higher result relevance as compared to string matching. This level of integration, with additional mapping of COI vocabularies to a hub taxonomy or thesaurus, could even provide automated link discovery between data assets in related, but autonomous COIs.[8]

Summary

This article describes a data sharing process with specific guidance for simple and advanced implementation scenarios. Components of this process have been implemented for enterprise sharing by the DoD Military Services. Much work remains in adoption and integration to enable DoDD 8320.02 net-centric sharing.

For Comments or Questions Contact:

Capt. Markyves Valentin, NGA/ASXT
(703-755-5411)
Markyves.J.Valentin@nga.mil

Dr. Lou York, MITRE
(781-271-3880)
lyork@mitre.org

[6] http://uddi.xml.org/ accessed 11/4/10.

[7] https://software.forge.mil/sf/sfmain/do/home accessed 11/4/10.

[8] "Metadata Mapper: The Other Resource Adapter", http://metadata.dod.mil/mdr/ns/geoint/news, Jan 2010.